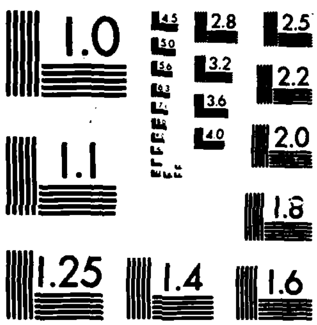
AD-A163 673 THE EFFECT OF VISUAL TASKLOAD ON CRITICAL FLICKER FREQUENCY (CFF) CHANGE D..(U) FEDERAL AVIATION ADMINISTRATION WASHINGTON DC OFFICE OF AVIAT.. 1/1

UNCLASSIFIED R I THACKRAY ET AL. OCT 85 FAA-AM-85-13 F/G 5/8 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DOT/FAA-AM-85-13

THE EFFECT OF VISUAL TASKLOAD ON CRITICAL FLICKER FREQUENCY (CFF) CHANGE DURING PERFORMANCE OF A COMPLEX MONITORING TASK

Richard I. Thackray and R. Mark Touchstone

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

OCTOBER 1985

Document is available to the public through the
National Technical Information Service,
Springfield, Virginia 22161.

Prepared for
U.S. DEPARTMENT OF TRANSPORTATION
Federal Aviation Administration
Office of Aviation Medicine
Washington, DC 20591

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA-AM-85-13	AD-A163673	

4. Title and Subtitle	
THE EFFECT OF VISUAL TASKLOAD ON CRITICAL FLICKER FREQUENCY (CFF) CHANGE DURING PERFORMANCE OF A COMPLEX MONITORING TASK	5. Report Date OCTOBER 1985
	6. Performing Organization Code
7. Author(s) Richard I. Thackray and R. Mark Touchstone	8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P. O. Box 25082 Oklahoma City, Oklahoma 73125	10. Work Unit No. (TRAIS)
	11. Contract or Grant No.
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, SW. Washington, DC 20591	
	14. Sponsoring Agency Code

15. Supplementary Notes

Work was performed under tasks AM-C-82/83/84/85-PSY-90

16. Abstract

The present study examined the effect of differing levels of visual taskload on critical flicker frequency (CFF) change during performance of a complex monitoring task. The task employed was designed to functionally simulate the general task characteristics of future, highly automated air traffic control systems in which passive monitoring is likely to be a principal job requirement. Forty subjects, divided into two equal-size groups, monitored displays containing either 8 or 16 alphanumeric targets. Nine critical events were randomly presented during each half-hour of the single 2-h session to which each subject was exposed. CFF thresholds were obtained prior to and following the sessions. Subjects monitored for the occurrence of two types of critical events. The first type consisted of a readily detectable change in an alphanumeric data block; the second kind of event was the occurrence of two aircraft (alphanumeric targets) at the same altitude on the same flight path. The results revealed that the more readily detectable critical events showed no evidence of performance decrement at either level of visual taskload. For the more difficult task of detecting critical altitude events, both CFF and performance showed evidences of fatigue that were confined entirely to the higher taskload condition. The findings are discussed with reference to fatigue and monitoring loads in highly automated air traffic control system concepts.

17. Key Words		18. Distribution Statement
Attention	Automation	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.
Monitoring	Performance	
Vigilance		
Air Traffic Control		
Critical Flicker Frequency		

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	18	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENT

The authors extend their sincere appreciation to Ms. Anna Chomiak of Oklahoma State University for the many hours and specialized assistance she provided in developing the multidimensional fatigue scale used in this study.

THE EFFECT OF VISUAL TASKLOAD ON CRITICAL FLICKER FREQUENCY (CFF) CHANGE DURING PERFORMANCE OF A COMPLEX MONITORING TASK

INTRODUCTION

With the development of increasingly automated air traffic control (ATC) systems, more and more of the functions currently performed by controllers will be assumed by the computer. As the controller's role gradually shifts from that of an active participant in control decisions to that of a monitor or overseer of a computerized operation, it seems likely that an increasing portion of the future controller's time will be spent in simply maintaining an awareness of actions being taken by the system and in verifying that those options presented by the computer are appropriate (Lipps et al. 1983; Swedish 1983). Current concepts of ATC automation, however, do not relieve the controller of his/her responsibility to adequately back up the system in the event of malfunctions or deviations from normal system operation. Even under the highest levels of automation, such as the AERA (Automated En Route Air Traffic Control) 3 concept, in which the controller is essentially out of the control loop, he/she still has the responsibility to maintain active awareness of system operation in order to act as a backup, if necessary. This is stated rather emphatically in the following quote pertaining to a fully automated concept: "Since the controller no longer plays an active role in the implementation of the control plan but still is responsible for the actions taken, it is critical that the controller be able to follow the control activities performed by the automation system so that the controller can intervene if necessary and make informed control decisions" (Lipps et al. 1983). If we assume that intervention might be a relatively infrequent occurrence in highly automated systems, questions relating to the maintainence of controller attentiveness, involvement, and readiness to react in emergencies while monitoring a largely computerized operation will become increasingly important considerations in the systems being planned.

In a previous study, we examined the effect that differing levels of visual taskload might have on the ability to sustain attention to a rather passive ATC monitoring task (Thackray et al. 1979). In this study, subjects were required to monitor either 4, 8, or 16 targets in order to detect occasional critical changes in the alphanumerics. While there was no evidence of performance decrement in the two lower taskload conditions, subjects assigned to the 16-target condition showed a significant decline in attention as reflected in a progressive increase in detection times over the 2-hour session. In a somewhat similar study, Howell et al. (1966) likewise found a direct relationship between the magnitude of vigilance decrement and the number of targets that subjects were required to search. Detailed analyses of data in both studies revealed that this increase in detection time under conditions of high target density was the result of a progressive increase in the duration of long detection times; shortest times remained relatively constant over successive periods of the monitoring session. Since the appearance of long response times, "blocks," or gaps during sustained performance of continuous or repetitive-type tasks is generally considered to be one of the more sensitive and acceptable criteria of fatigue (Bertelson and Joffe 1963), it would seem reasonable to hypothesize that a requirement to continuously monitor a large number of targets over a prolonged period of time demands considerable effort, and that the

increase in frequency of long response times found in our study was a manifestation of the fatigue resulting from this effort. A similar hypothesis was advanced by Howell et al. (1966) to account for the performance decrement found in their study.

Even though a fatigue explanation would appear to be one of the more plausible ones to account for the performance decrement that occurred only under the highest taskload condition, Thackray et al. (1979) were generally unsuccessful in finding ancillary evidence to support this. Except for an apparent increase in the frequency of brief eye closures that occurred primarily in the 16-target condition, most of the physiological and subjective measures of arousal or fatigue employed showed changes during task performance that were no greater in the 16- than in the 4- or 8-target conditions. Although it is believed that the obtained decrement in performance was indeed a manifestation of fatigue, it would clearly be desirable to obtain alternative indices, apart from performance, that could be used to detect its occurrence. Such "indirect" indicators of the presence of fatigue might then be used under operational conditions in which direct assessment of performance change could be difficult, if not impossible.

Within recent years, there has been renewed interest in the use of critical flicker frequency (CFF) as a measure of fatigue, in particular the mental fatigue associated with performance of repetitive or vigilance-type tasks demanding sustained attention over prolonged periods of time (Davies et al. 1983, Grandjean 1979). CFF, which may be defined as "...that rate of successive light flashes from a stationary light source at which the sensation of flicker disappears and the light becomes 'steady'" (Simonson and Brozek 1952), is a measure that appears to reflect the number of impulses capable of being processed by the retinal-cortical system per unit time (Levander and Lagergren 1973). Grandjean and his colleagues have done most of the recent work with this measure and were quite successful in using CFF as a measure of fatigue among air traffic controllers (Grandjean et al. 1971). In their study of Swiss controllers at the Zurich airport, CFF was found to decline gradually over the first 6 hours of a 10-hour work period and then to decline more abruptly during the final 4 hours. A parallel pattern of change was noted for subjective ratings of tiredness.

Similar changes in CFF, although generally of lower magnitude, have also been reported in laboratory studies of mental fatigue. Thus, Baschera and Grandjean (1979) compared CFF change during the performance of three versions of a repetitive task; the versions all required the same motor responses but differed in cognitive/perceptual demand. CFF was found to decline significantly in both the least demanding (most monotonous) and the most demanding versions, but not in the moderately difficult one. This finding was in agreement with their hypothesis that both the "monotonie" of a repetitive, low difficulty task and the "fatigue" resulting from performance of a high difficulty, perceptually demanding task would produce common CFF central nervous system changes. Grandjean (1979), in fact, has stated that boredom and "mental fatigue" produce symptoms that are nearly the same and that it may be impossible to distinguish between the two.

In a subsequent study, Weber et al. (1980) examined the psychophysiological and subjective changes associated with performance of four versions of a somewhat similar task, all versions of which required the same motor activity but differed in the level of perceptual discrimination needed. CFF was found to decline significantly over time in all versions of the task, but unlike the previous findings of Baschera and Grandjean (1979), only the most demanding task version showed a decline that was significantly greater than that shown by the other three. Subjective fatigue patterns differed among the versions, with drowsiness and boredom being greatest in the two versions requiring least perceptual discrimination, and tension characterizing the two most difficult ones.

The above studies, and others reviewed by Davies et al. (1983), all suggest that continued performance of repetitive tasks demanding sustained attention is accompanied by a decline in CFF. However, laboratory findings differ with regard to the relationship of this decline to taskload. Thus, while Weber et al. (1980) found that only the heaviest taskload produced a CFF decline that was significantly greater than that of the other three taskload conditions, Baschera and Grandjean (1979) found that performance of a low difficulty task resulted in a CFF decline that was equivalent to that produced by a much more demanding task. It is evident that, if CFF is to be of any value as a measure of fatigue in ATC applications, a clearer understanding of its relationship to taskload is required.

The primary purpose of the present study, then, was to examine the effect of two levels of visual taskload on CFF change during the performance of a simulated air traffic control monitoring task. In addition, the study also sought to determine whether a multidimensional scale of subjective fatigue, modeled after similar scales devised by Wolf (1967) and Kinsman et al. (1973), might yield information on specific patterns of fatigue, unique to the high and low taskload conditions and not apparent through the use of CFF alone. This latter possibility, that high and low taskload conditions may elicit qualitatively different patterns of subjective fatigue, has been suggested by Wolf and reflects a growing belief that multidimensional scales encompassing independent clusters of subjective symptoms may reveal unique patterns of fatigue associated with specific tasks and task conditions that cannot be revealed by simple unidimensional scales alone (Kinsman and Weiser 1976).

The task employed represented an updated version of the radar monitoring task used in our laboratory over approximately the last 10 years. The previous version, which initially used film strips and later a computer-generated display, was designed to functionally simulate the general task characteristics of future, highly automated ATC systems in which passive monitoring is expected to be a principal task requirement. While this earlier version proved to be useful in the investigation of a variety of variables affecting complex monitoring performance, it had certain limitations that restricted its use to particular kinds of research questions. For example, it, like virtually all other existing vigilance tasks, was developed primarily for the purpose of investigating factors affecting the detection of only one type or class of signal over time. As investigators are beginning to recognize, modern operational vigilance tasks, such as those involving the monitoring of automated processes, involve more than simply detecting and responding to infrequent

changes in unidimensional stimuli. They frequently involve complex multidimensional discriminations (Mackie 1984) in which stimulus detection or identification may be followed by interpretation of significance, decisions as to appropriate action, implementation of actions, and evaluation of consequences (Craig 1984). Considerations such as these led us to the decision to develop a new version of our task that would allow us to investigate the effects of prolonged monitoring, not only on perceptual processes, but on a range of cognitive behaviors as well. In addition, it was felt that any revision of our radar monitoring task should provide for much greater flexibility of person-computer interaction in order to allow for studies dealing with optimal allocation of function in automated ATC systems.

The task described in the present report represents the initial version of our updated radar monitoring task that, with subsequent modifications as needed, will evolve into a testbed for evaluating a variety of questions related to ATC automation. The ultimate goal is the development of a task that will allow us to examine separately most of the major behaviors that will probably continue to be required of controllers, although probably less often in the case of some of these behaviors, even under moderately high levels of automation. These behaviors include the detection of both simple and complex critical events (target changes), short-term memory, decision making, action selection, and action verification. Only data pertaining to the detection of critical events, however, are reported in the present study. The total task is still in the process of development, and data relevant to the other subtask elements are being analyzed in the context of this development process and will be reported on later in a separate study. For the purpose of providing the reader with details regarding the total task performed by the subjects, all aspects of the task are described in the procedure.

METHOD

Subjects. Thirty-two male and eight female paid university students volunteered to participate in the study. They were randomly assigned, in equal male-female proportions, to one of two target density (8- or 16-target) groups. Subjects ranged in age from 18 to 29 years, and none had prior experience with the task used or previous ATC training. All had 20/20 vision, corrected or uncorrected, and all had no reported hearing loss.

Radar Simulation Task. The basic experimental equipment consisted of a Digital Equipment Corportion (DEC) VS11 19-in (49-cm) graphics display, keyboard, and joystick, all of which were interfaced with a VAX 11/730 computer (DEC). The computer was used both to generate input to the display and to process responses of the subjects. The VS11 was incorporated into a console designed to closely resemble an ATC radar unit. Two diagonal, nonintersecting flight paths were located on the display, along which aircraft targets could move in either direction. A given aircraft's location was displayed as a small "blip" on the flight path, and an adjacent alphanumeric data block identified the aircraft and gave its altitude and groundspeed. Aircraft were updated as to location and any change in alphanumerics in a continuous, clockwise manner, such that a complete update occurred every 6 s. The overall visual impression was that of a series of discrete jumps flowing in a circular pattern. This movement pattern approximates very closely the way in which aircraft targets are updated in

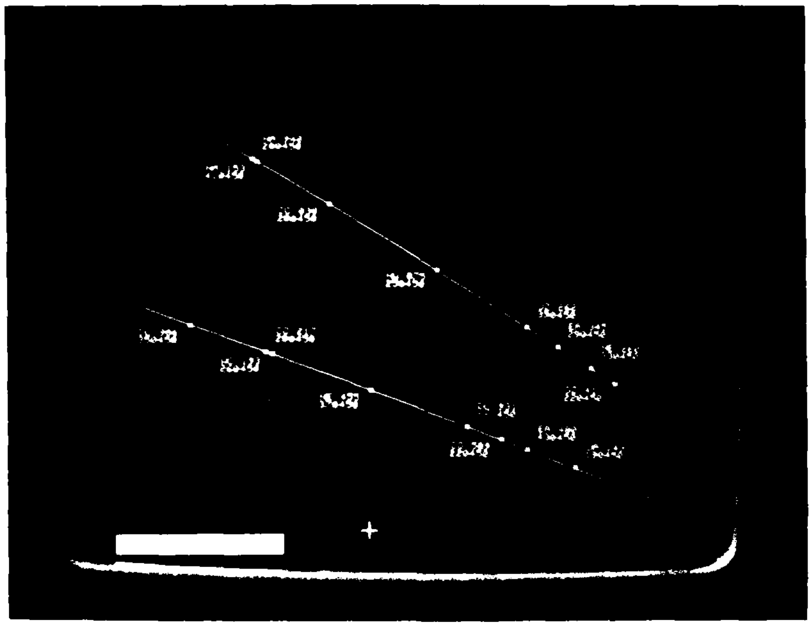


Figure 1. A typical target configuration as displayed to the subject.

Figure 2. The simulated ATC work station. Only the console on the left was used in this study.

contemporary ATC radars with computer-generated alphanumeric displays. Figure 1 shows a typical target pattern as displayed to the subject, with the total console-display configuration shown in Figure 2.

The subject's task was to continually monitor the display for the occurrence of one of two types of designated changes, referred to as critical events, in the alphanumeric data blocks. Stimulus duration for these critical events was always 90 s; if a subject failed to detect a critical event within this 90-s period, the data block containing the change reverted to its previous state.

The first type of critical event was a readily detectable one and consisted of the appearance of three X's in place of the three altitude numbers in a given data block. Subjects were told that this replacement of an altitude value by the three X's signified that some malfunction had occurred resulting in a loss of altitude information. Upon detection of such an event, subjects were told to press a button on the console labeled "XXX malfunction," move a joystick-controlled cursor over the data block containing the critical event, and press another button on the joystick control unit. This last response "corrected" the malfunction by replacing the three X's with the previous altitude value. The second type of critical event was the occurrence of two aircraft at the same altitude on the same flight path. As soon as such an event was noted, subjects pressed a second console button labeled "Altitude Check." It was next determined whether the two aircraft were moving towards each other, moving away from each other, or traveling in the same direction. On the basis of this determination, subjects then pressed either a "Conflict" or a "No Conflict" button. (All aircraft in this simulation were assigned a speed of 450 mph. Thus, aircraft could not overtake one another, and only targets moving towards each other would constitute a potential conflict situation.) Following a "conflict" decision, the cursor was positioned over one of the two conflicting aircraft and the joystick control button was pressed. This caused a new altitude value to appear in the lower left of the screen that, subjects were told, represented a value selected by the computer to resolve the conflict. Subjects then verified that the computer-assigned altitude did not result in a conflict with some other aircraft on the flight path. If no new conflict was created, a keyboard entry was made that assigned the new altitude value to one of the two previously conflicting aircraft. (Under the simulation used, a newly assigned altitude never conflicted with the altitude of any other aircraft.)

Whenever a "no conflict" response was made, no further action ensued, since no change in altitude was required. Subjects were told that the altitude of one of the two nonconflicting aircraft would eventually change to some other value (this time was variable, but always less than the 90-s stimulus duration period) and that they had to remember that they had responded to this particular pair of aircraft. If they failed to remember and responded a second time, an error was recorded.

For a given taskload condition (8 or 16 targets), the number of targets on each flight path was kept equal at all times; as one left the screen, another appeared. In order to prevent any occurrence of data block overlap, two constraints were necessary: (a) speed was held constant at 450 miles per hour for all aircraft, and (b) all data blocks for targets moving from left to right were positioned above the flight path, while those moving right to left were

located below. Nine critical events occurred in each 30-min period, with no more than one event present at any given time. Of these nine events, three were XXX's, three were conflicting altitude changes, and three were nonconflicting changes. These events were arranged in a quasi-random order with the restriction that each of the three types of events had to occur at least once in both the first and second 15 min of each 30-min period. Subjects were given no information regarding the frequency of events or their order of occurrence. The times between events (interstimulus intervals) ranged from 126 to 302 s with a mean of 200 s.

CFF Measurement. A Lafayette Instrument Company model 12025 Flicker Fusion Apparatus with binocular viewing chamber served as the basic equipment for CFF measurement. A circular target area 0.5 inches in diameter was viewed monocularly at a distance of 14.5 in, subtending a visual angle of 2 degrees, 10 min. Two Sylvania R1166 glow modulator tubes were located in the viewing chamber behind milk glass diffusion filters; the left one served as the stimulus light source, while the right one drove a phototransistor that was connected directly to the digital inputs of the computer. A light/dark ratio of 1:1 was used for the stimulus light, and the intensity of the test patch was 5 mL.

The CFF apparatus was modified so that thresholds were obtained directly by subjects themselves. The psychophysical method employed is referred to as the Bekesy tracking procedure and, although commonly used in audiometry, has not to our knowledge been used previously to measure CFF. (Recently, Ginsburg and Cannon (1983) and Kohfeld (1985) have reported using this technique to measure visual contrast sensitivity.) A stepping motor, mechanically attached to the frequency control of the CFF apparatus, allowed subjects to increase or decrease the frequency of the visual stimulus by means of a hand-held microswitch. With the stepping motor in operation, rate of change in the flickering light was approximately 0.43 Hz/s. Subjects were instructed to press the microswitch whenever the light appeared to be flickering and to continue to hold the switch down until fusion occurred, whereupon they were to release the switch until a flickering sensation was just noticed. They were told to repeat this sequence until the light in the viewing chamber disappeared, signaling a rest period. CFF thresholds were determined from two sessions; one occurred at the beginning of the experiment and the other at the end. The sessions consisted of five 40-s trials, with each trial followed by a 40-s rest period. A small red indicator light alerted subjects at the beginning of each trial. All sessions were begun with flicker rate initially set at 35 Hz.

A computer program was developed that determined, within each 40-s trial period, the flicker frequency of the visual stimulus at the time each button press and release occurred. The values thus obtained gave the upper and lower threshold frequencies for each ascending and descending series. CFF was defined as the average of the upper and lower values obtained within each 40-s trial period.

The CFF apparatus and computer were located in an adjacent room from which the subject was monitored via closed-circuit TV. Indirect lighting was used in the subject's room, and the level of illumination at the display was 5.6 lux.

Multidimensional Scale. The scale employed was derived from an earlier pilot study in which 35 fatigue-related adjectives were rated by 71 undergraduate

during the course of normal, everyday activities. Adjectives used in developing the scale were those that appeared related to the fatigue factors (exhaustion, drowsiness, and tension/nervousness) isolated earlier by Wolf (1967), or that seemed related to a fatigue dimension that we labeled task aversion/boredom. This latter dimension was included because the items selected (e.g., bored, frustrated, apathetic) appeared to be associated with the feeling of conflict that Bartley and Chute (1947) regard as a necessary precursor to feelings of fatigue. A varimax rotation of a principal components analysis of the scores obtained in the pilot study yielded four interpretable factors. The items loading on each of these factors along with factor loadings are shown in Table 1. The only factor that does not appear clearly distinct from the others is the one labeled general activation. This factor is similar to one of the factors in Thayer's (1978) Activation-Deactivation scale of arousal and seemingly represents the opposite pole of the sleepy/exhausted factor. Since the factors contained differing numbers of items, the factor scores used in the present study consisted of the means of the item scores rather than the sums.

Table 1. Subjective fatigue factors and their item loadings.

Sleepy/ Exhausted		Nervous/ Tense		Task Aversion/ Boredom		General Activation	
Items	Loadings	Items	Loadings	Items	Loadings	Items	Loadings
Tired	.70	Relaxed	-.56	Bored	.77	Vigorous	.50
Fatigued	.80	Leisurely	-.52	Frustrated	.35	Activated	.75
Rested	-.74	Anxious	.70	Empty	.33	Peppy	.74
Lazy	.42	Peaceful	-.43	Negative	.44	Energetic	.73
Physically tired	.80	Irritable	.56	Annoyed	.53	Lively	.58
Aching	.53	Nervous	.66	Monotony	.50		
Drowsy	.63	Jumpy	.62	Dull	.67		
Exhausted	.81	Calm	-.67	Uninterested	.72		
Mentally sluggish	.50	Tense	.61	Apathetic	.72		
Weary	.61			Unwilling	.52		
Sleepy	.62						
Worn out	.75						

Procedure. On arrival, the subject was taken to the testing room, and an orientation tape was played. The orientation stated that this was one of a series of studies designed to investigate the role of the controller in increasingly automated ATC systems and that information their participation would provide might be used in planning the job/task design of future systems. Following this, subjects completed the multidimensional scale of subjective fatigue, rating items in accordance with how they felt at that time.

After completing the fatigue scale, subjects listened to the taped instructions for CFF measurement. The instructions indicated that any time the visual stimulus appeared to be flickering or fluttering to them, they were to press down on the response button until the light appeared to be steady; at this point the button was to be released until a flickering or fluttering sensation again occurred. This sequence of pressing and releasing the button was to be continued as long as the light in the viewing chamber remained on; when the light went off, it signaled the beginning of a rest period. It was emphasized that the button should be pressed down immediately at the first appearance of flicker and released just as soon as it appeared steady. They were also instructed to try to keep their eyes looking directly at the light at all times during a test trial. Five 40-s trials separated by 40-s rest periods were then administered.

Immediately following CFF measurement, subjects were given task instructions and separate practice in responding to each of the three kinds of critical events. This was followed by a practice session in which the various kinds of critical events were presented in a random order. Twenty-one critical events (seven of each kind) occurred during the practice sessions, which lasted a total of 21 min. On rare occasions, additional practice was given if the subject appeared to have difficulty with any of the procedures.

The experimental session lasted 2 hours. In order to add a greater element of realism to the task, a tape recording of background noises recorded in actual air traffic control radar rooms was played continuously during the 2-hour task session. Sound level of this noise at the subject's head location was 62 dBA. It was not expected that this would have any effect on performance, since an earlier study, using the previous version of our monitoring task, failed to find any significant performance effects of this noise at a considerably higher (80 dBA) level (Thackray 1982).

At the completion of the 2-hour task period, five CFF trials were again administered along with the subjective rating scale. For half the subjects, CFF preceded the subjective scale, while this order was reversed for the remaining half.

RESULTS

Performance Data. As described earlier, two levels of stimulus difficulty were employed in this study. In the first level, subjects were required to simply scan the display for the occurrence of three X's that replaced a three-digit altitude value in one of the targets on the screen. The second level, which was more difficult, required subjects to make continual comparisons of each target's altitude with the altitude values of all other targets on a given flight path in order to detect the occasional occurrence of two targets at the same altitude. These two levels of stimulus difficulty will henceforth be referred to as the low difficulty (LD) and high difficulty (HD) levels respectively.

Figure 3 shows mean detection times across 30-min periods for both types of critical events under 8- and 16-target taskload conditions. Analyses of variance applied to the LD data revealed a significant main effect for target density ($F(1/38)=17.71$, $p<.01$) and a significant target density by periods

interaction ($F(3/114)=3.15$, $p<.05$). Tests of simple effects of the interaction revealed the two taskload conditions to differ significantly during the first ($F(1/19)=10.06$, $p<.01$), second ($F(1/19)=18.24$, $p<.01$), and third ($F(1/19)=6.95$, $p<.05$) half-hour periods, but not during the fourth ($F(1/19)<1.00$). Additional tests of simple effects revealed the differences between periods to be significant for the 16-target condition ($F(3/114)=2.83$, $p<.05$), but not for the the 8-target condition ($F(3/114)=1.97$, $p>.10$).

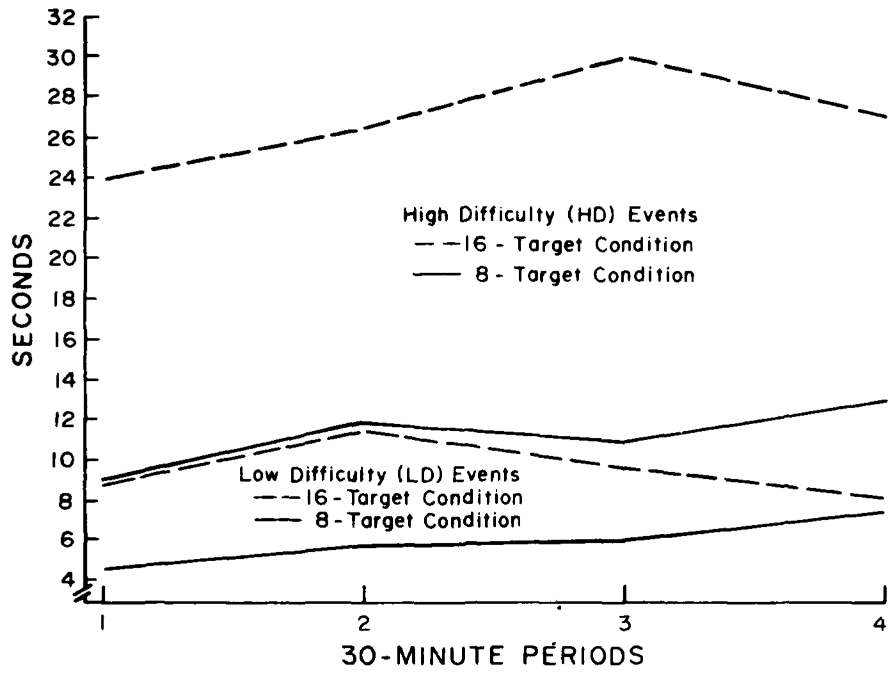


Figure 3. Mean detection times across 30-min periods for LD and HD events under the two taskload conditions.

Unlike the data obtained for LD events, the more difficult task of detecting targets at the same altitude does not appear differentially influenced by taskload conditions. Except for some reversals, the general pattern shown in both task conditions is that of an increase in detection time from the beginning to the end of the session. Analyses of variance supported this impression by revealing significant main effects for both target density ($F(1/38)=89.24$, $p<.001$) and periods ($F(2/114)=3.14$, $p<.05$), with no significant interaction.

With regard to errors of omission, the more readily detectable LD events were seldom missed, and there was no indication that the frequency with which these events were missed differed as a function of target density. Thus, no events were missed under the lower taskload (8-target) condition, with only two subjects missing one event each under the 16-target condition. HD events, however, were missed more frequently, and these omissions were definitely influenced by taskload. Thus, while only 1 subject missed a single HD event under the lower taskload condition, 15 (or 75 percent) of the subjects missed at

least one event under the 16-target condition. In addition, failures to detect HD events increased in frequency over the task session. Out of the 12 HD events that occurred during the first hour of the session, subjects missed, on the average, less than 1 in 12 ($Mn=0.6$). During the second hour, however, the average almost tripled to a mean of 1.6 events missed. A Wilcoxon test revealed this difference between first and second hours to be significant ($p<.05$).

CFF Data. CFF values for measurements obtained at the beginning and end of the experimental session for the 8- and 16-target conditions are shown in Figure 4. As is evident from the figure, while both groups appear to differ little at the

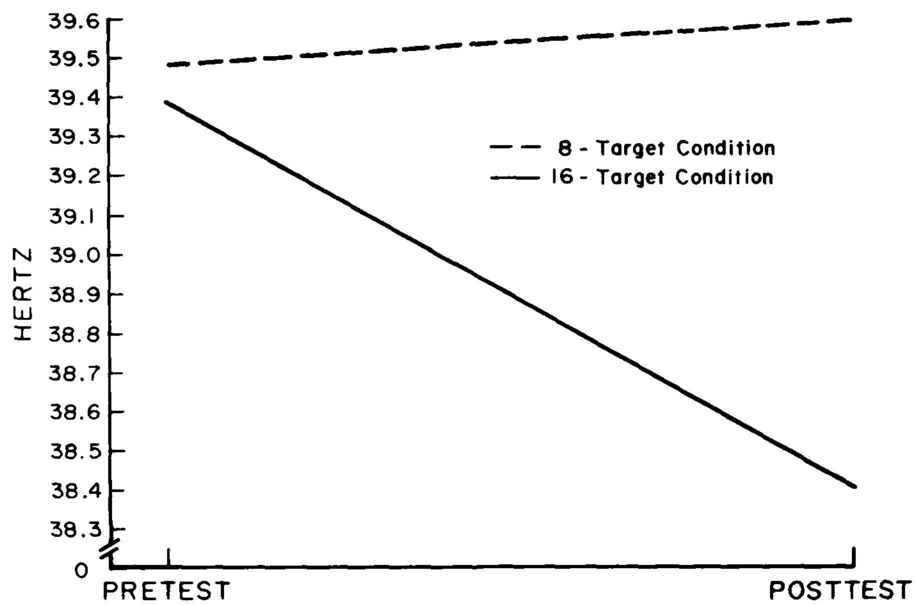


Figure 4. Pretask and posttask CFF values for the two taskload conditions.

beginning, there is marked divergence at the end of the session. As might be expected from the data shown in this figure, an analysis of variance revealed neither the main effect for groups, nor the main effect for measurement periods, to be significant ($p>.05$). There was, however, a significant interaction of groups by measurement periods ($F(1/38)=4.60$, $p<.05$). Tests of simple effects of the interaction revealed the pretask to posttask decrease in CFF under the 16-target condition to be significant ($F(1/38)=7.40$, $p<.05$), while the apparent increase in CFF under the 8-target condition was not ($F<1.00$).

Because of the significant decline in CFF found for the 16-target condition, further comparisons were made to determine whether magnitude of CFF change was related to performance change. Decrease in CFF from pretask to posttask measurement periods was correlated with increase, from the first to the second hour, in the number of HD events missed, and also with increase in mean detection time to HD events over these same two 1-hour periods. While the

correlation of CFF decrease with omission error increase failed to reach significance ($r=-.30$, $p>.05$), the relationship was in the expected direction. There was no evidence of any relationship between CFF change and the magnitude of increase in mean detection time to HD events ($r=.03$, $p>.05$).

Subjective Fatigue Data. The subjective fatigue scale was administered prior to and following the 2-hour task session. Table 2 shows mean scores for each of

Table 2. Mean pretask and posttask factor scores for the two taskload conditions.

	Sleepy/ Exhausted		Nervous/ Tense		Task Aversion/ Boredom		General Activation	
Taskload	Pre	Post	Pre	Post	Pre	Post	Pre	Post
8 target	3.84	2.79	3.27	3.53	4.45	3.48	2.57	3.56
16 target	4.15	2.96	3.57	3.61	4.48	3.54	2.63	3.59

the four factors under the two taskload conditions. (The lower the score, the greater the feeling state represented by the factor.) Since scores could range from 1 (definite presence of the feeling state) to 5 (definite absence of the feeling state), it is evident that most of the factor scores shown in the table suggest moderate, rather than extreme, levels of feeling. Analyses of variance revealed significant main effects for pretask to posttask change for the sleepy/exhausted factor ($F(1/38)=80.08$, $p<.001$), the task aversion/boredom factor ($F(1/38)=67.18$, $p<.001$), and the general activation factor ($F(1/38)=37.70$, $p<.001$). None of the other main effects or interactions of the analyses conducted on the above factors was significant, nor were any effects significant for the nervous/tense factor. Thus, while performance of the radar monitoring task significantly increased feelings associated with the sleepy/exhausted and task aversion/boredom factors and decreased feelings of general activation, there were no differences in the magnitude of these changes that could be attributed to the two taskload conditions. Nor were any differences at all associated with the nervous/tense factor.

Because the earlier scale developed by Wolf had separate factors for exhaustion and drowsiness while the present scale combined these two, the Wolf scale was reconstructed from comparable items in order to determine whether, by separating the two factors, relationships with taskload might be found that would be more in accord with Wolf's predictions. Thus, tired, lazy, drowsy, mentally sluggish, and sleepy were items included in the present scale that were essentially the same as those loading on Wolf's drowsiness factor, while physically tired, aching, exhausted, and worn out were items that loaded on his exhaustion factor. Means of the item scores comprising each factor are shown in Table 3. Analyses of variance conducted on these data revealed significant main

Table 3. Mean pretask and posttask scores for Wolf's drowsiness and exhaustion factors under the two taskload conditions.

	Drowsiness		Exhaustion	
Taskload	Pre	Post	Pre	Post
8 target	3.97	2.58	4.01	3.36
16 target	4.27	2.71	4.17	3.51

effects for factors ($F(1/38)=16.49$, $p<.001$), for order of administration ($F(1/38)=69.98$, $p<.001$), and a significant order by factor interaction ($F(1/38)=31.62$, $p<.001$). No other effects were significant. Although no differences as a function of target density were found in this analysis, the significant order by factor interaction, coupled with an examination of Table 3, clearly reveals that it was an increase in feelings of drowsiness rather than exhaustion that was produced by the 2-hour monitoring session.

DISCUSSION

The results of the present study revealed a rather mixed pattern of performance change. Detection times for readily detectable stimuli (LD events) showed a slight, nonsignificant increase over the 2-hour session under the lower taskload condition and a pattern of increasing followed by decreasing detection times under the higher taskload condition. These events were virtually never missed under either target density condition. For the more difficult task of detecting two aircraft at the same altitude (HD events), detection times increased significantly during the session under the higher taskload condition, but also increased significantly, and in a parallel manner, under the lower condition. While none of the above data would support any suggestion of greater impairment under the 16-target condition, this was not the case with data regarding failures to detect HD events. It will be recalled that, when an altitude event occurred, resulting in two aircraft at the same altitude, this event remained on the screen for a period of 90 s. If the event was not responded to within this time period, it reverted back to its previous value, and the failure to respond was recorded as a missed stimulus. The number of times that these altitude events were completely missed increased significantly during the task session, but only when 16 targets were being monitored; no increase occurred when monitoring was confined to just eight targets.

The data for missed events suggest a type of performance decrement in the higher taskload condition that was absent in the lower condition. However, it must be recognized that this apparent difference between conditions could simply be an artifact resulting from the need to place some limit on the length of time that a critical event remained on the screen. If it is assumed that each of the HD events missed under the 16-target condition would eventually have been detected

had the events not been "removed" after the 90-s stimulus duration period, one could then view these missed events simply as long detection times that overlapped the 90-s timeout period with increasing frequency as the session progressed. If this assumption is correct, then a comparable increase in long detection times may also have occurred under the 8-target condition but, because of the lower mean detection times associated with this condition, even extreme latencies would likely have had values that fell considerably below the 90-s stimulus duration period. In order to examine the comparability of the distributions of long detection times under the two taskload conditions, each subject's longest detection time for HD events in each of the half-hour periods of the session was obtained and means and standard deviations computed. These data are shown in Table 4. An analysis of variance revealed a significant

Table 4. Means and standard deviations (in seconds) of maximum detection times over half-hour periods for the two taskload conditions.

	Periods			
Taskload	1	2	3	4
8 Target	14.43 (4.71)	19.07 (7.27)	18.62 (8.45)	21.52 (15.79)
16 Target	39.29 (12.54)	39.49 (16.28)	46.26 (12.33)	41.72 (12.01)

conditions effect ($F(1/38)=106.01$, $p<.001$), but neither the periods effect nor the conditions by periods interaction was significant ($p>.10$). The lack of a significant periods effect might be expected if missed events under the 16-target condition did indeed represent long detection times that were not included in the latency distributions; the lack of a significant interaction effect suggests that no real increase in long detection times occurred during the session under the 8-target condition.

The increase in frequency of missed HD events under the 16-target condition did not appear to be due simply to a decline in scanning activity. This was determined from data obtained for the LD events. In the context of the present study, one of the reasons for including these events was to provide a measure of detection time to readily detectable stimulus changes, which in turn served as an index for assessing the frequency and adequacy of scanning behavior during the course of the monitoring session. Thus, if failures to scan the screen under the 16-target condition occurred more frequently during the second hour, one might reasonably expect to see an increase in detection time to LD events and possibly even missed stimuli. However, detection times to these stimuli under the higher taskload condition were virtually the same at the beginning and end of the session, and there was no evidence of any increase in the already extremely low frequency of missed LD events. (Only two were missed, and both of these occurred during the first hour.)

If we compare the findings of this study with those of our earlier one in which we compared monitoring performance to 4, 8, and 16 targets (Thackray et al. 1979), a number of similarities become apparent. Both studies found evidence of performance decline occurring over a 2-hour session of continuous monitoring, but only under the highest taskload condition. In our previous study, this decrement was due primarily to an increase in the duration of long detection times, and this, in turn, was taken to be an indication of increasingly frequent lapses or fluctuations of attention. The task employed in this earlier study required only the detection of a fixed, three-digit altitude number; these events were almost never missed within the 60-s stimulus duration period. In the present study, it was an increase in the frequency of HD events missed, rather than an increase in time to detect such events, that appeared to characterize the 16-target condition. If, as indicated previously, one makes the reasonable assumption that each of these events would have been detected, albeit with extremely long detection times, had the events not been "removed" after the 90-s stimulus duration period, then the findings of the two studies become surprisingly similar. For reasons that are not entirely clear, both studies found evidence of apparent lapses in the reception or processing of sensory information that occurred when monitoring 16 targets, but not when 8 or fewer targets were monitored. In the earlier study, these "lapses" were manifested as long detection times, while, in the present one, they took the form of missed events. In neither study was there any indication that higher visual taskloads resulted in performance impairment simply because of a reduction in scanning activity.

We initially hypothesized that apparent lapses in the reception or processing of sensory information under conditions of high monitoring load were likely to be manifestations of some fatigue process. This hypothesis was supported by the findings of the present study that CFF declined significantly (approximately 1 Hz) under the 16-target condition but showed no change when only 8 targets were monitored. A decline of this magnitude would appear to be within the range of values reported in other studies of CFF change with fatigue (Grandjean et al. 1977). These findings with respect to CFF, then, would suggest that this measure might be useful in operational situations to assess levels of fatigue. Clearly, Grandjean and his colleagues (Grandjean et al. 1971) found this to be the case in their study of fatigue among Swiss air traffic controllers. It should be noted, however, that that study also found that changes in subjective measures of "tiredness" paralleled the decline in CFF. In the present study, none of the subjective measures of fatigue that were employed showed any differences, either qualitative or quantitative, as a function of taskload. A similar lack of relationship of subjective fatigue to taskload was also found in our earlier study (Thackray et al. 1979). While one might speculate almost endlessly as to the reasons why subjective fatigue showed no relationship to taskload in these two investigations, one reasonable possibility is suggested by the findings of the Grandjean et al. (1971) study of controller fatigue. In that study, most of the physiological and subjective indices of fatigue showed only a slight to moderate decrease (increase in fatigue) during the first 6 hours of the 10-hour work period; the greatest decrease did not occur until the final 4 hours. On the basis of those findings, it is conceivable that the lack of relationship of subjective fatigue to taskload that was found in both the present study and in our earlier one (Thackray et al. 1979) as well, may have been due to the duration of the monitoring sessions employed. Had sessions

lasting longer than 2 hours been used, significant relationships of subjective fatigue to taskload might have been obtained.

CONCLUSIONS

As noted earlier, the increasing automation of air traffic control will likely result in the controller becoming less of an active participant in control decisions and more of a passive monitor of a system largely run by computers. It is frequently assumed that this apparent reduction in workload will enable the controller to monitor a larger number of aircraft and hence increase productivity (Swedish 1983). For reasons that are not yet well understood, in both the present study and our earlier one, and using quite different display configurations and task designs, performance decrement, seemingly resulting from some fatigue process, was found only under the highest monitoring loads employed; there was little or no evidence of any decline in performance under the lower taskloads used in each study. What was particularly noteworthy was the finding in the present study of a significant increase under the higher taskload condition in the number of potential conflicts that were simply not "seen" during the 90-s period they were on the screen. While it must be recognized that these results were obtained under simulated ATC conditions, and that some degree of caution should be exercised in generalizing to operational tasks of the future, the findings at least suggest that the assumption of increasing productivity through an increase in the number of aircraft handled per controller may need to be considered rather carefully in highly automated ATC concepts where passive monitoring is a major task component.

It is evident that the present research has raised certain questions requiring further study. These questions include, but are not limited to, the following:

(1) Since the present study found evidence of performance decrement while monitoring 16 but not while monitoring 8 targets, at what level of visual taskload would evidence of a significant decline in performance begin to occur?

(2) What variables, apart from target density, may have contributed to the decrement found under the 16-target condition? While considerable information exists with regard to variables influencing performance decrement in simple vigilance tasks, relatively little research has been conducted to determine whether these same variables affect behavior in a similar manner in complex monitoring tasks requiring visual search.

(3) Why did measures of subjective fatigue fail to show any differences as a function of taskload?

(4) How does the apparent fatigue that is associated with monitoring relatively large numbers of targets affect other relevant ATC behaviors (e.g., decision making, short-term memory, action selection) that were measured but not analyzed in the present study?

Some or all of these questions will be addressed in a planned study dealing with the differential effects of fatigue on the separate behaviors involved in ATC conflict detection and resolution.

References

BARTLEY, S. H. and CHUTE, E., 1947, Fatigue and Impairment in Man (New York: MCGRAW-HILL).

BASCHERA, P. and GRANDJEAN, E., 1979, Effects of repetitive tasks with different degrees of difficulty on critical fusion frequency (CFF) and subjective state. Ergonomics, 22, 377-385.

BERTELSON, R. and JOFFE, R., 1963, Blockings in prolonged serial responding. Ergonomics, 6, 109-116.

CRAIG, A., 1984, Human engineering: The control of vigilance. In Sustained Attention in Human Performance (New York: WILEY), pp. 247-291.

DAVIES, D. R., SHACKLETON, V. J., and PARASURAMAN, R., 1983, Monotony and boredom. In Stress and Fatigue in Human Performance (Edited by R. HOCKEY) (New York: WILEY).

GINSBURG, A. P. and CANNON, M. W., 1983, Comparison of three methods for rapid determination of threshold contrast sensitivity. Investigative Opthalmology and Visual Science, 24, 798-802.

GRANDJEAN, E., 1979, Fatigue in industry. British Journal of Industrial Medicine, 36, 175-186.

GRANDJEAN, E., BASCHERA, P., MARTIN, E., and WEBER, A., 1977, The effects of various conditions on subjective states and critical flicker frequency. In Vigilance: Theory, Operational Performance, and Physiological Correlates (Edited by R. R. MACKIE) (New York: PLENUM).

GRANDJEAN, E. P., WOTZKA, G. SCHAAD, R., and GILGEN, A., 1971, Fatigue and stress in air traffic controllers. Ergonomics, 14, 159-165.

HOWELL, W. C., JOHNSTON, W. A., and GOLDSTEIN, I. L., 1966, Complex monitoring and its relation to the classical problem of vigilance. Organizational Behavior and Human Performance, 1, 129-150.

KINSMAN, R. A. and WEISER, P. C., 1976, Subjective symptomology during work and fatigue. In Psychological Aspects and Physiological Correlates of Work and Fatigue (Edited by E. SIMONSON and P. C. WEISER) (Springfield, Illinois: CHARLES C. THOMAS), pp. 336-405.

KINSMAN, R. A., WEISER, P.C., and STAMPER, D. A., 1973, Multidimensional analysis of subjective symptomology during prolonged strenuous exercise. Ergonomics, 16, 211-226.

KOHFELD, D. L., 1985, An evaluation of visual contrast sensitivity measures. In Trends in Ergonomics/Human Factors II (Edited by R. E. EBERTS and C. G. EBERTS) (North-Holland: ELSEVIER), pp. 83-92.

LEVANDER, S. E. and LAGERGREN, K., 1973, Four vigilance indicators for use with a microcomputer. Reports from the Psychological Laboratories, University of Stockholm, No. 381.

LIPPS, A. W., SWEDISH, W. J., and ZIMMERMAN, B. C., 1983, Operation and functional description of the AERA packages. U. S. Department of Transportation, Federal Aviation Administration Report No. DOT/FAA/ES-83-10, Washington, D. C.

MACKIE, R. R., 1984, Research relevance and the information glut. In Human Factors Review: 1984 (Edited by F. A. MUCKLER) (Santa Monica, California: THE HUMAN FACTORS SOCIETY, INC.), pp. 1-11.

SIMONSON, E. and BROZEK, J., 1952, Flicker fusion frequency. Background and applications. Physiological Review, 32, 349-378.

SWEDISH, W. J., 1983, Evolution of advanced ATC automation functions. Mitre Working Paper Report 83W149, The Mitre Corporation, McLean, Virginia.

THACKRAY, R. I., 1982, Some effects of noise on monitoring performance and physiological response. Academic Psychology Bulletin, 4, 73-81.

THACKRAY, R. I., BAILEY, J. P., and TOUCHSTONE, R. M., 1979, The effect of increased monitoring load on vigilance performance using a simulated radar display. Ergonomics, 22, 529-539.

THAYER, R. E., 1978, Factor analytic and reliability studies on the activation-deactivation check list. Psychological Reports, 42, 747-756.

WEBER, A., FUSSLER, C., O'HANLON, J. F., GIERER, R., and GRANDJEAN, E., 1980, Psychophysiological effects of repetitive tasks. Ergonomics, 23, 1033-1046.

WOLF, G., 1967, Construct validation of measures of experiential fatigue. Perceptual and Motor Skills, 24, 1067-1076.

END

FILMED

3-86

DTIC